Logistic Regression

Raguvir Kunani

Discussion 10

November 6, 2019

Motivating Logistic Regression

We've seen linear regression, where we predict a response variable that can take on any scalar value from a set of explanatory variables or features.

But we also want to be able to predict things that are of the Yes/No form as well. That's where logistic regression comes in.

Why can't we use linear regression to predict a Yes/No (i.e. 0 or 1)?

Linear Regression gives us a line, which means we will have predictions outside the range $[0, 1]$.

The Logistic Regression Model

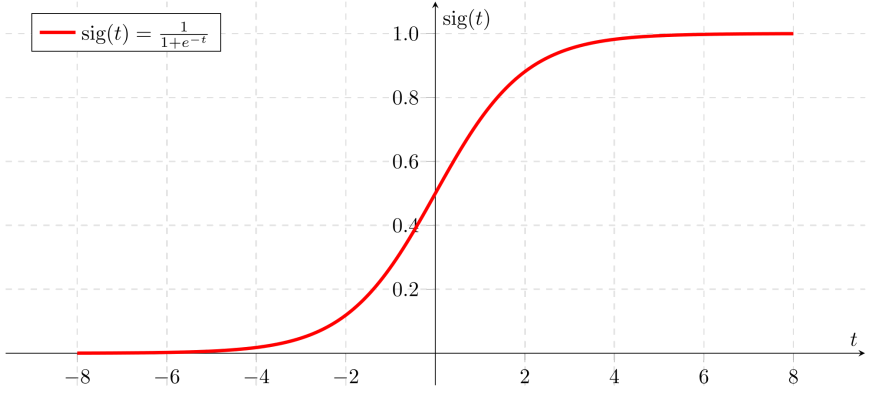
Here is our logistic regression model:

$$P(Y = 1|x) = \sigma(x^T \beta)$$

linear regression

> If the math ever gets too much for you, just take a step back and remember that we are just trying to come up with probabilities.

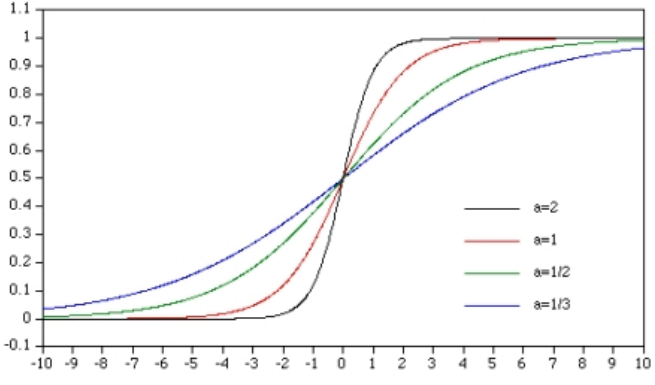
The σ represents the sigmoid function, which looks like this:



THIS SLIDE ISN'T SUPER IMPORTANT, it just provides a pictorial view of what logistic regression is doing

A Closer Look at the Sigmoid Function

The sigmoid function is really doing all the magic for us, so it's worth taking a closer look at it.



Notice that different choices for the input to the sigmoid (labeled a in the diagram) result in different shapes of the sigmoid. Why is this good for prediction? Different choices β give sigmoids of different "steepness"; logistic regression finds the steepest curve that fits the data

$f_\beta(x_i) = \sigma(x_i^T \beta)$

How Do We Choose β?

Same as before! We will minimize a loss function. This time, it's not MSE though. Here's our new loss function, called cross-entropy loss:

y_i: true values (0 or 1)

$$L(\beta, X, y) = \frac{1}{n} \sum_{i=1}^{n} (-y_i \log(f_\beta(X_i)) - (1 - y_i) \log(1 - f_\beta(X_i)))$$

I don't know about you, but I don't want to solve for the β that minimizes this. Instead, we use gradient descent (more on this in another slide).

Using the Model to Make Predictions

Recall our logistic regression model is

$$P(Y = 1|x) = \sigma(x^T \beta)$$

The sigmoid will output a number between 0 and 1, which we then need to convert into a prediction. To do this, we just choose a threshold (usually 0.5) and predict a $Y = 1$ if $\sigma(x^T \beta)$ is above the threshold and predict $Y = 0$ if $\sigma(x^T \beta)$ is below the threshold.

Worksheet!

Feedback Form

This *anonymous* form is for me to learn what I can do to ensure you all get the most of discussion and lab. This form will be open all semester, and I'll be checking it regularly. Be as ruthless as you want, I promise my feelings won't get hurt.

Feedback Form: tinyurl.com/raguvirTAfeedback